2019 DOE Vehicle Technologies Office
Annual Merit Review

HIGH-PERFORMANCE COMPUTING (HPC) AND BIG DATA SOLUTIONS FOR MOBILITY DESIGN AND PLANNING

Jane Macfarlane
Lawrence Berkeley National Laboratory
June 12, 2019

Project ID: eems037

BERKELEY LAB
Argonne NATIONAL LABORATORY
Pacific Northwest NATIONAL LABORATORY

OVERVIEW

TIMELINE

- Start: October 2017
- End: September 2020
- 50% complete

BUDGET

- Total project funding
- $6M / 3 years
- $2M per year / 3 Labs

PARTNERS

- CalTrans Connected Corridors
- HERE Technologies

BARRIERS

- Metropolitan scale networks are too complex to model in reasonable compute time.
- Sensors for capturing dynamics provide limited view and are difficult to mine for relevant information.
- Optimization of energy, travel time and mobility across complex networks has yet to be accomplished for real-world metropolitan scale networks.

RELEVANCE – PROJECT OBJECTIVES

- Overall Objective:
  - Develop HPC tools to **rapidly model large scale transportation networks** using real-world, near real-time data. Integrate energy, travel time and mobility measures to determine optimization opportunities.
- Objectives this Period:
  - Improved capability for capturing metropolitan scale traffic dynamics with **dynamic routing capabilities** – the first step to modeling dynamics with active control.
  - Improve **estimates of the energy cost** and productivity loss of congestion **using data-driven approach**.
  - Analyze real-world sensor data to understand network demand and **improve link level models** in the simulation.
- Impact:
  - Develop new **active control ideas for connected vehicles** that will optimize energy, travel time and mobility for normal traffic conditions and networks under stress.

PROJECT GOALS

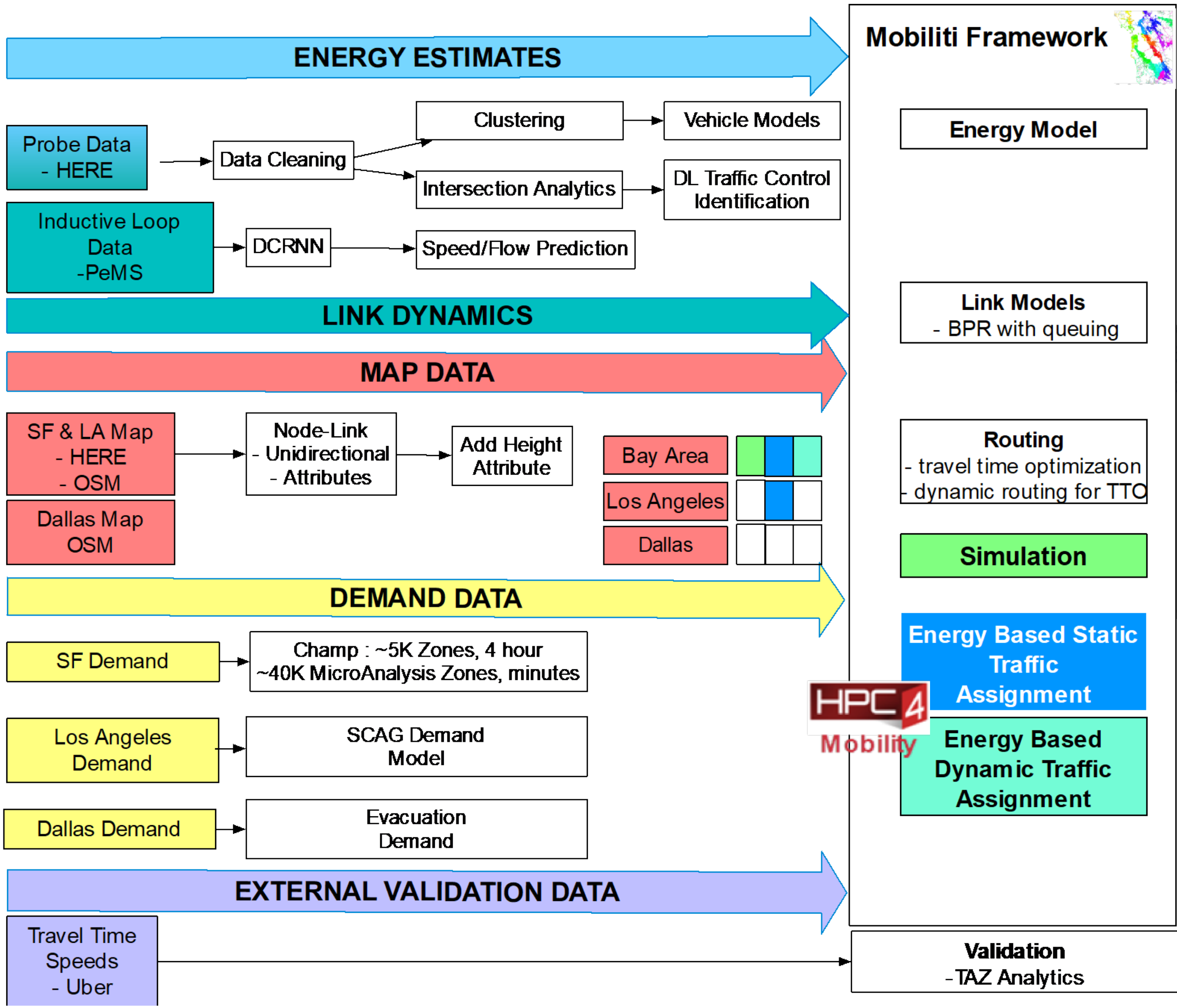
- Develop the data science and an HPC supported computational framework for building next-generation transportation /mobility and operational analytics.
  - Define appropriate role of HPC, ML and big data analytics in transportation problem domain.
    - Identify key transportation specific HPC technology gaps.
    - Define and access real-world datasets.
  - Automate the collection and validation of real-world transportation data.
    - Define data veracity analytics for real-world data.
    - Evaluate ML as a mechanism for analyzing geospatial data.
    - Automated ingestion with coupled analytics and ML models.
  - Develop asynchronous distributed state HPC transportation network models.
    - Develop ML models for estimating energy use.
    - Develop HPC enabled for transportation network modeling tool for large-scale networks.
    - Integrate HPC traffic assignment models for contrasting optimization modeling.
  - Couple automated data ingestion, learning systems and large-scale modeling in a full metropolitan scale modeling platform.
    - Energy tracing through HPC enabled models
    - Define tool integration path and architecture.

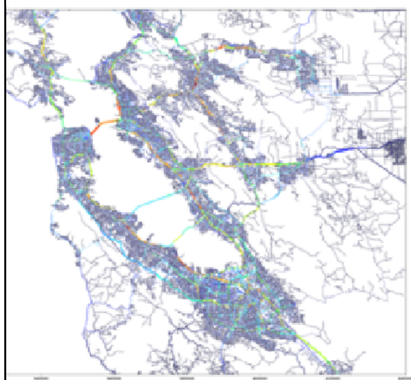
KEY MILESTONES

Define appropriate role of HPC, ML and big data analytics in transportation problem domain.	Defined goals for developing metropolitan scale modeling. Alliances with SF and San Jose.	Identified traffic assignment optimization research for integration. Collaboration established with Dallas Ft Worth/TTI.		Continuing
Automate the collection and validation of real-world transportation data.		Go/NoGo - Demonstrated good modeling of speed and flow with DCRNN with automated ingestion of loop detectors.	Use of probe data as virtual sensors to augment current loop detectors geospatial range.	On Track
		Developed data driven ML models for estimating energy consumption.	Integrated energy estimation.	On Track
Develop large-scale HPC enabled transportation network models.	Go/NoGo - Mobiliti model developed that models metropolitan scale network with compute time < 1 minute.	Go/NoGo - Mobiliti model developed that models metropolitan scale network capability of dynamic routing.	Investigate Active Control methods focused on reduction of energy and increased mobility.	On Track
Couple automated data ingestion, learning systems and large-scale modeling in a full metropolitan scale modeling platform.			Go/NoGo - Integration of ML models into the link dynamic models in Mobiliti.	

Key Go/NoGo milestones have been achieved in FY17/FY18

COMPONENTS OF THE APPROACH

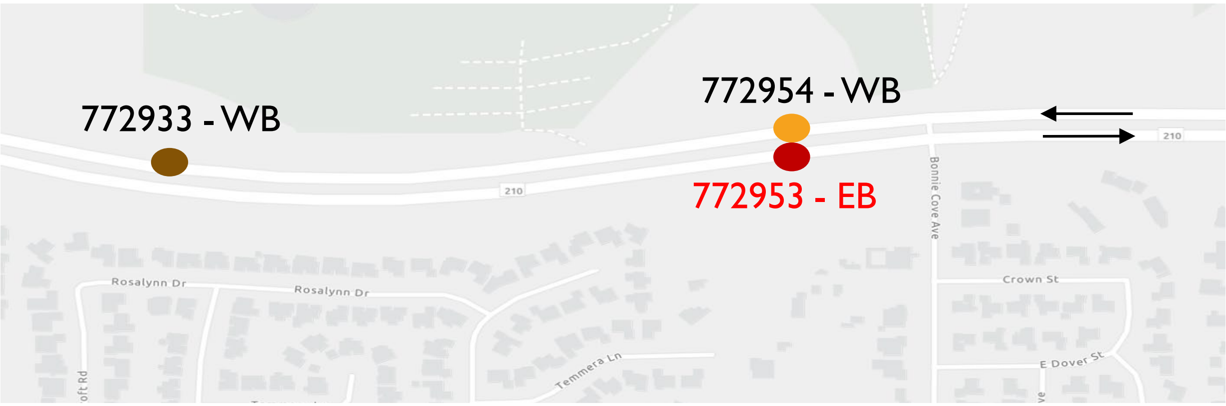


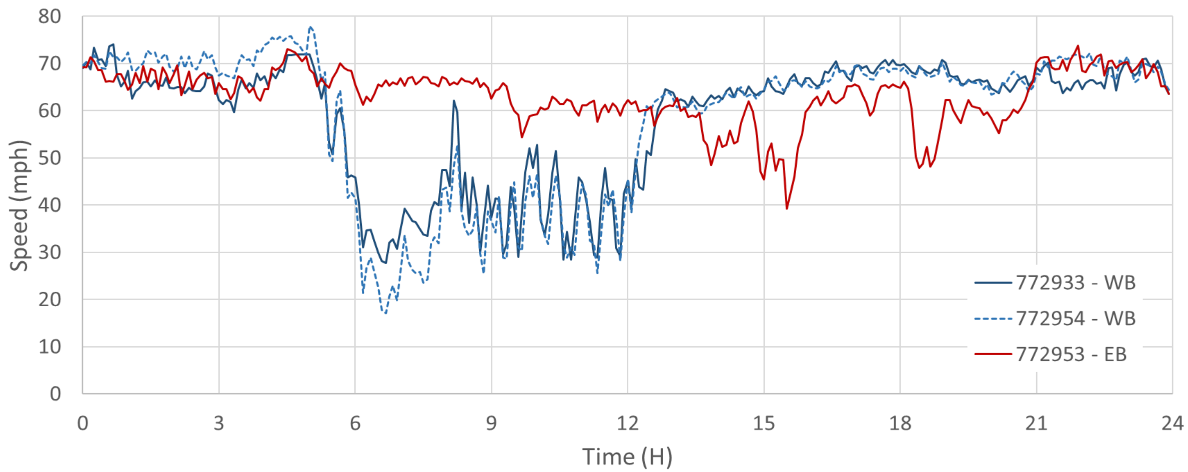


Metropolitan Scale Traffic Dynamics

CHALLENGES WITH SENSOR DATA MODELING

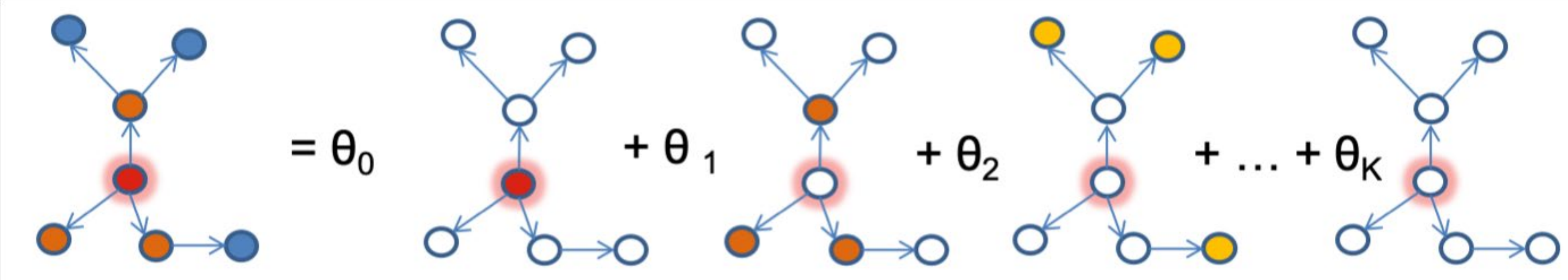
PeMS Data : Inductive loop sensors in major highways

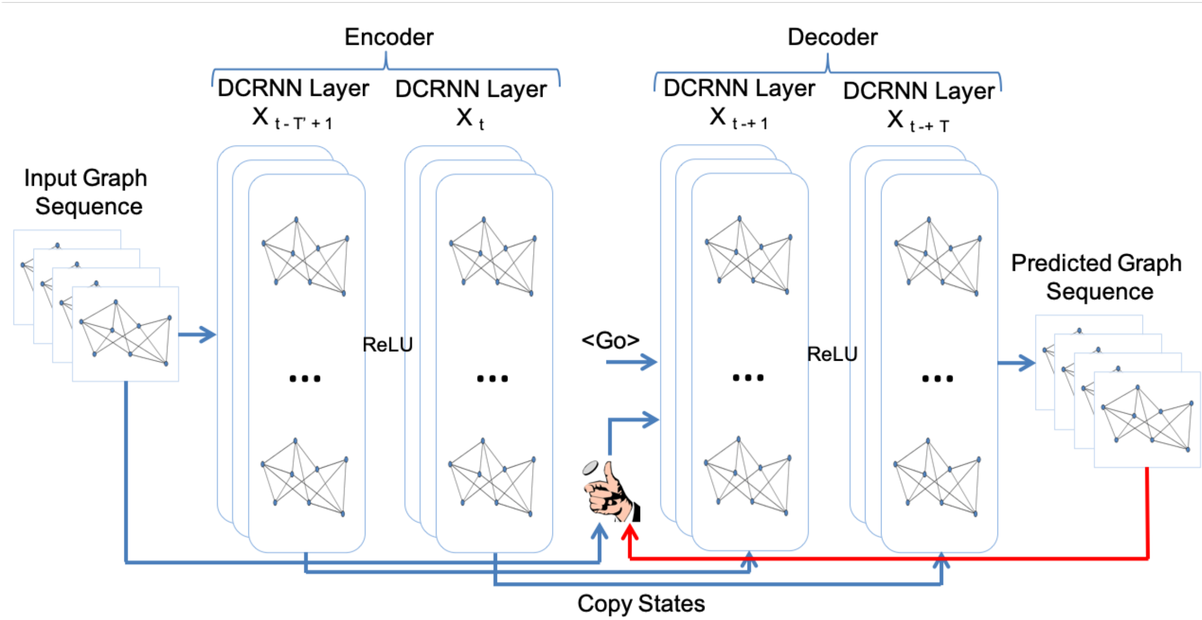




- Complex spatial dependency
- Non-stationary temporal dynamics
- Non-Euclidean spatial geometry
- Modelling each sensor independently fails to capture the spatial correlation

FORECASTING VEHICLE DYNAMICS USING DCRNN

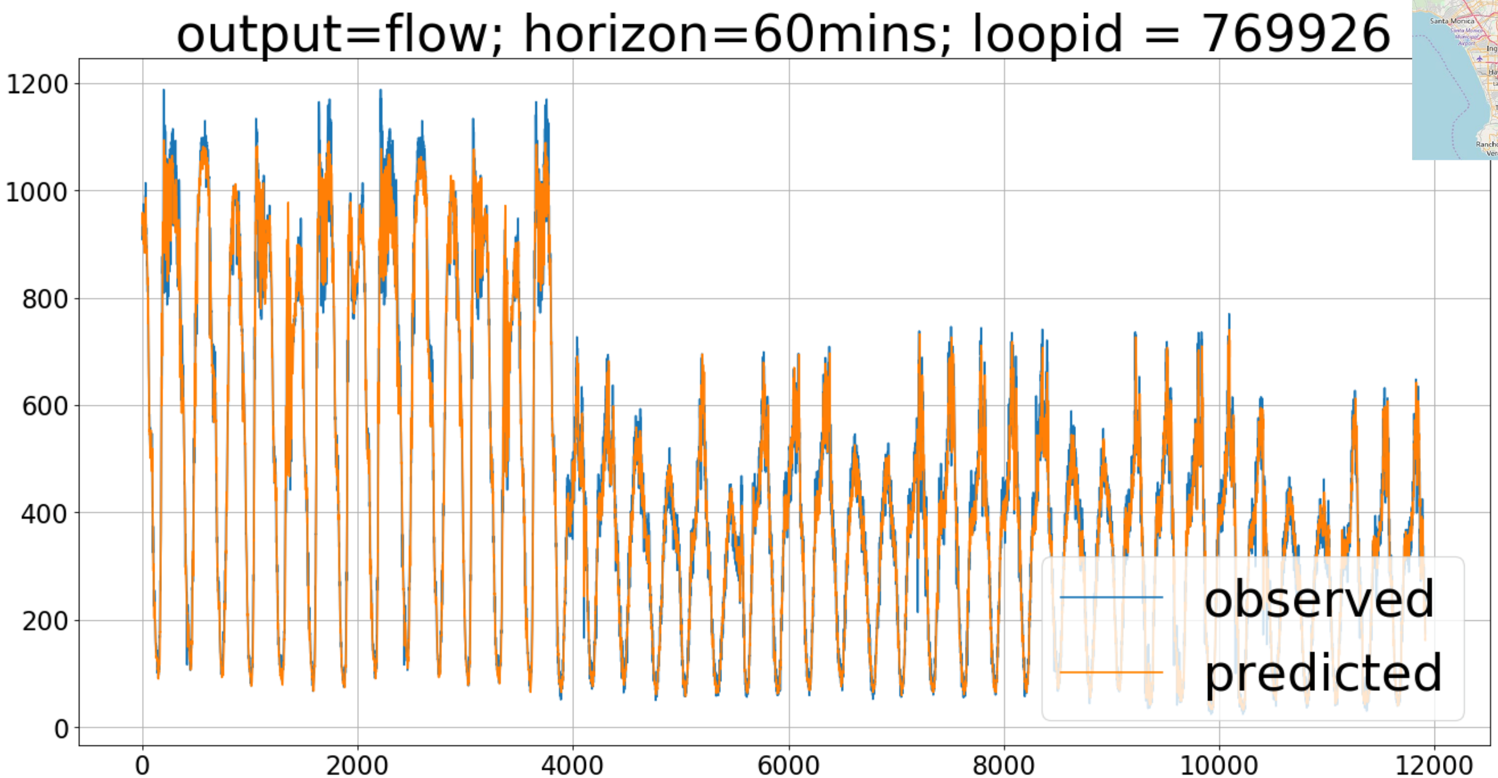


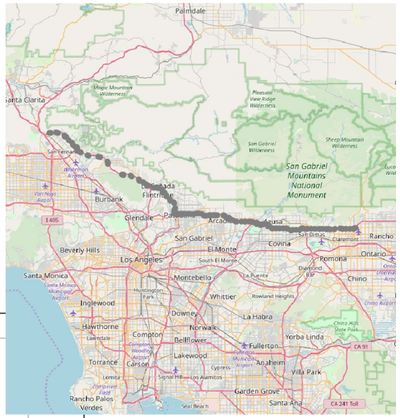


Combining the Diffusion Convolution with a Recurrent Neural Network into a Diffusion Convolutional Recurrent Neural Network (DCRNN) allows for predicting speeds and flows from inductive loop sensors.

FLOW PREDICTION : 162 LOOP DETECTORS

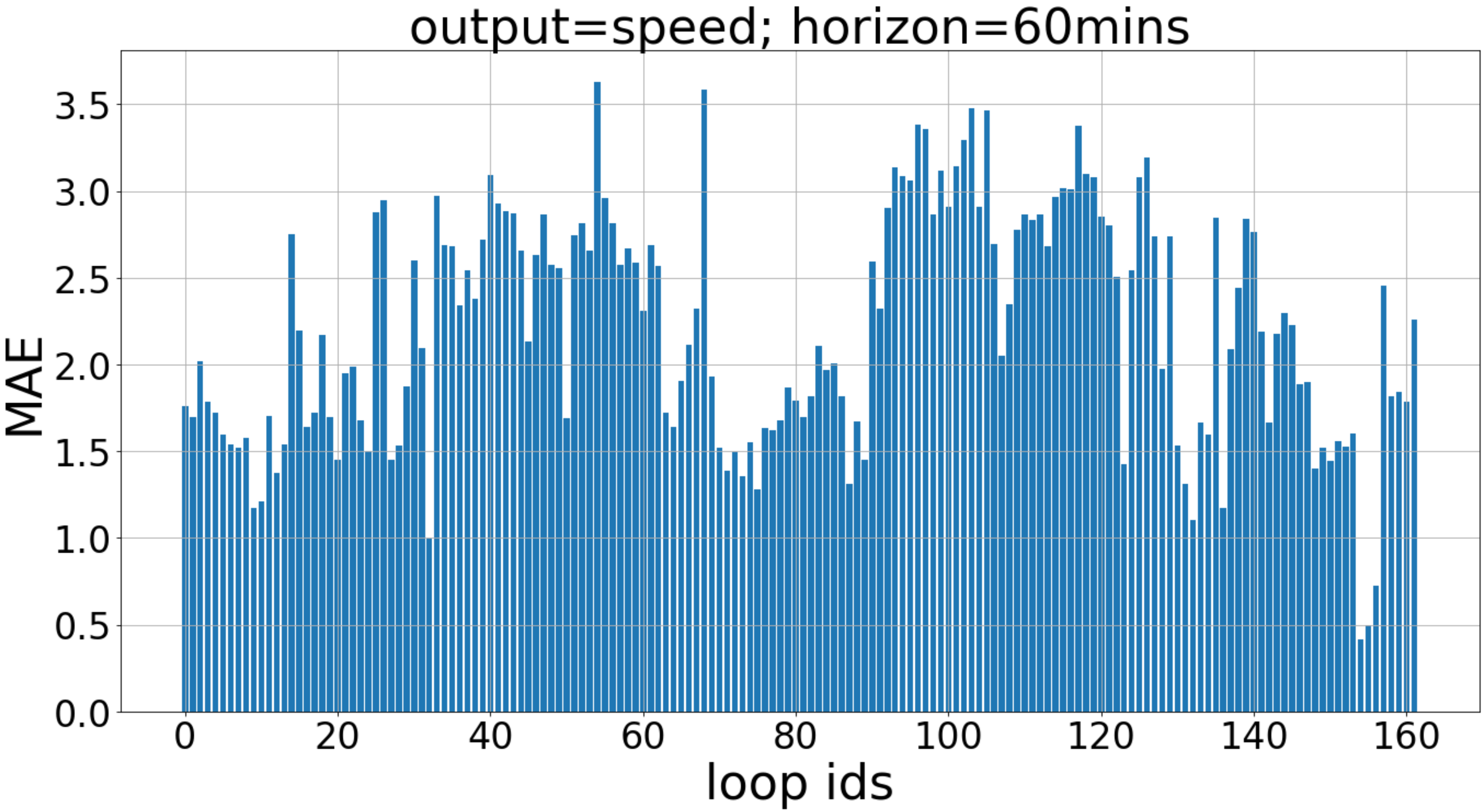
- District: Los Angeles (D7)





DCRNN tracks the real-world flows

MEAN AVERAGE ERROR FOR ALL LOOP DETECTORS



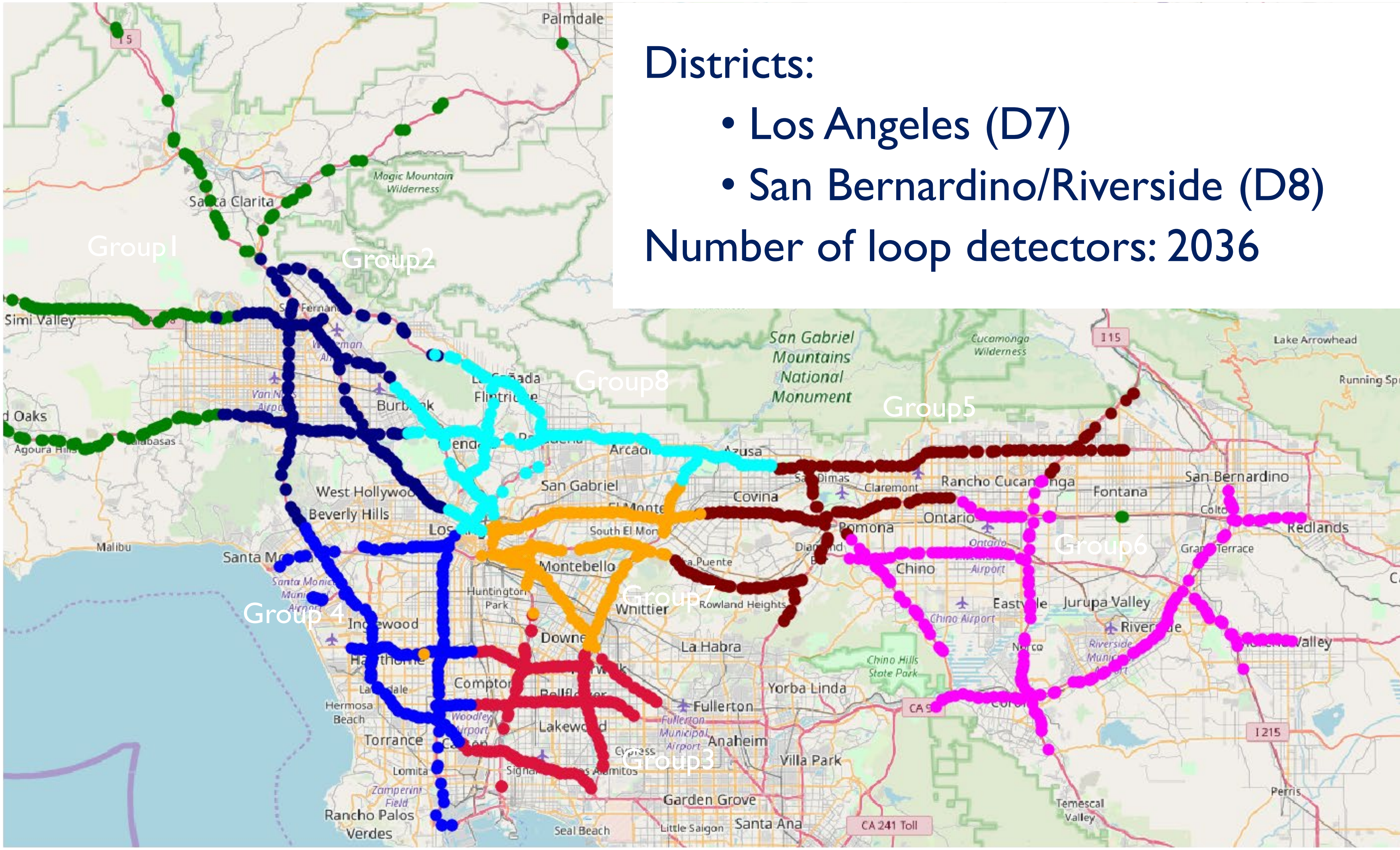
MAE is usually under 3 miles per hour

MAP PARTITIONING USING METIS

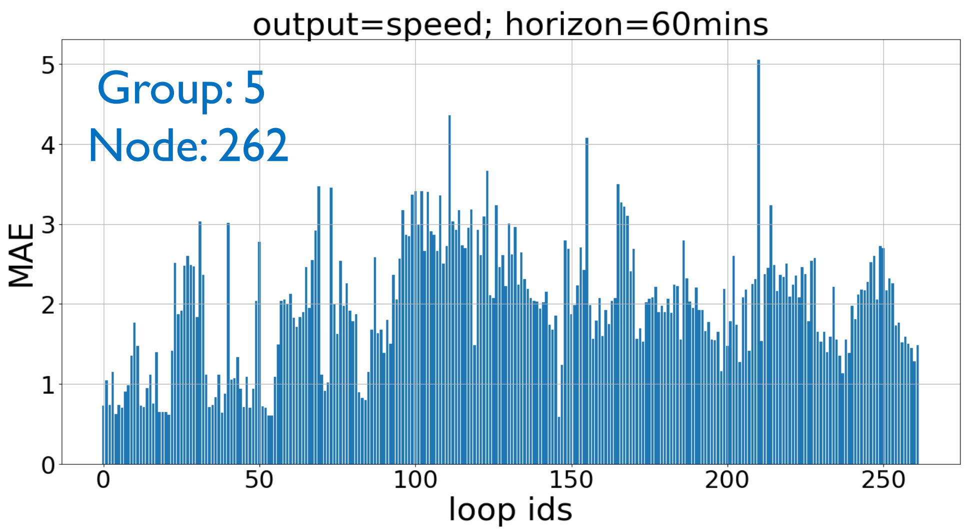
Districts:

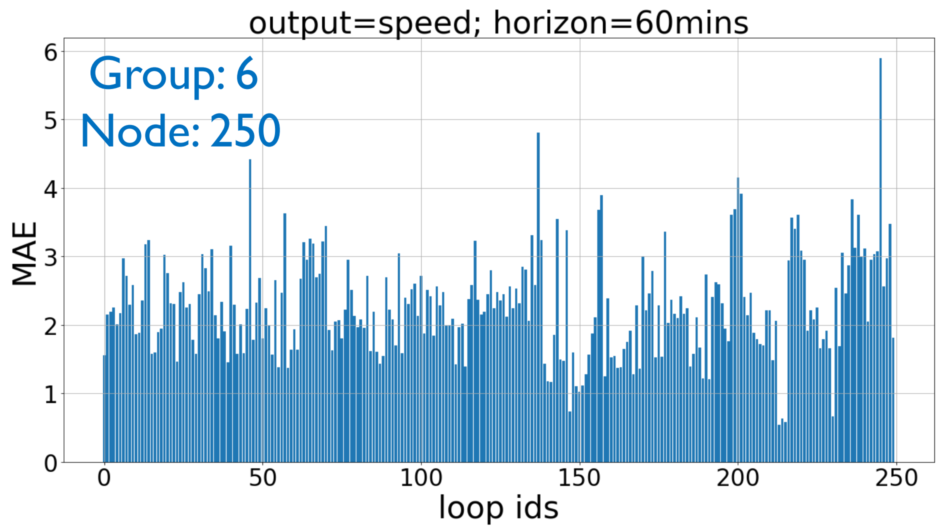
- Los Angeles (D7)
- San Bernardino/Riverside (D8)

Number of loop detectors: 2036

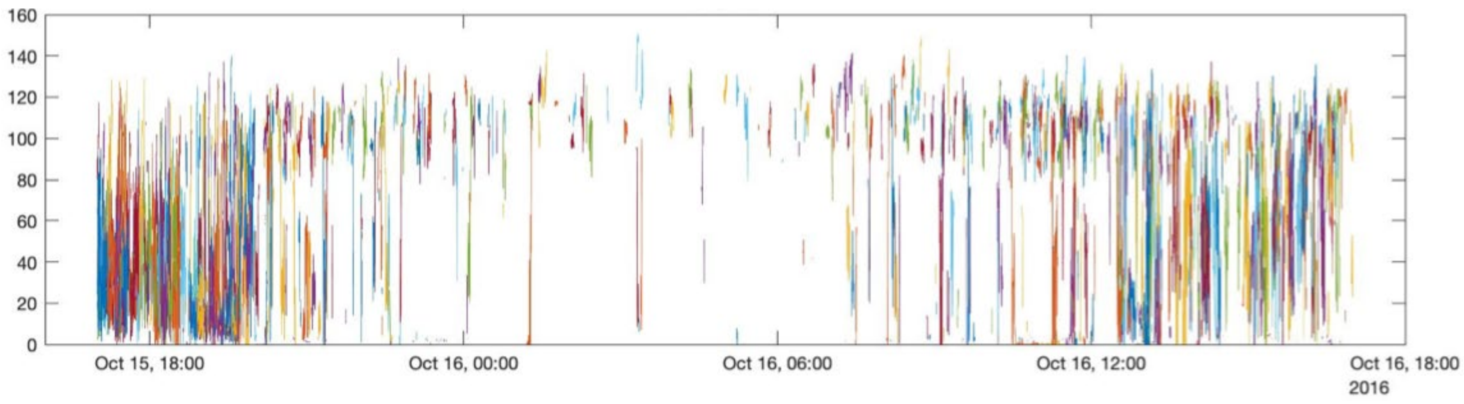


DCRNN RESULTS : NEXT STEP MOBILE DEVICE INTEGRATION





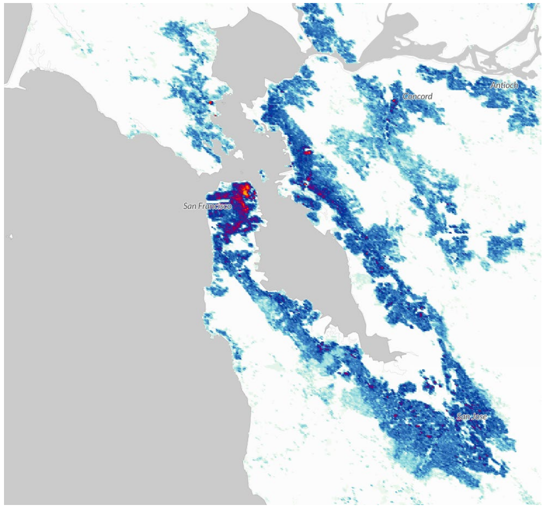
MAE is usually under 3 miles per hour

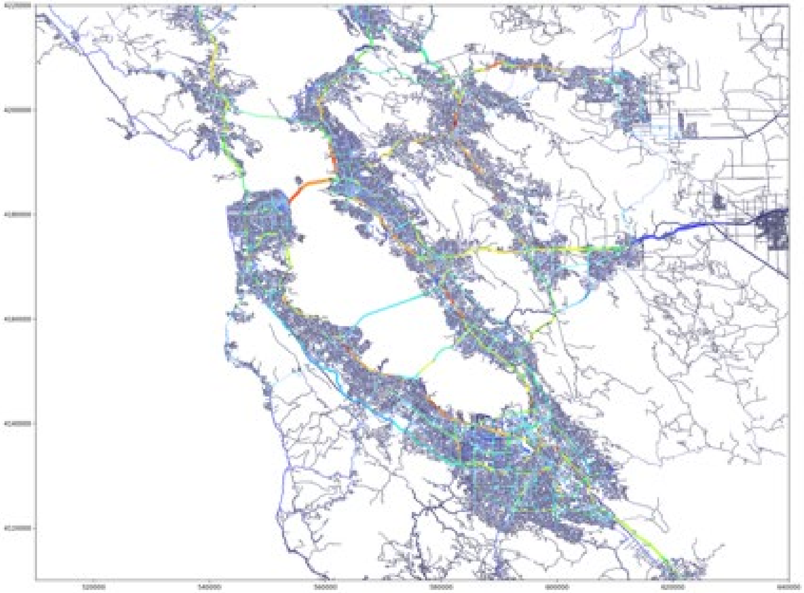


Mobile device trajectories for I210 segment

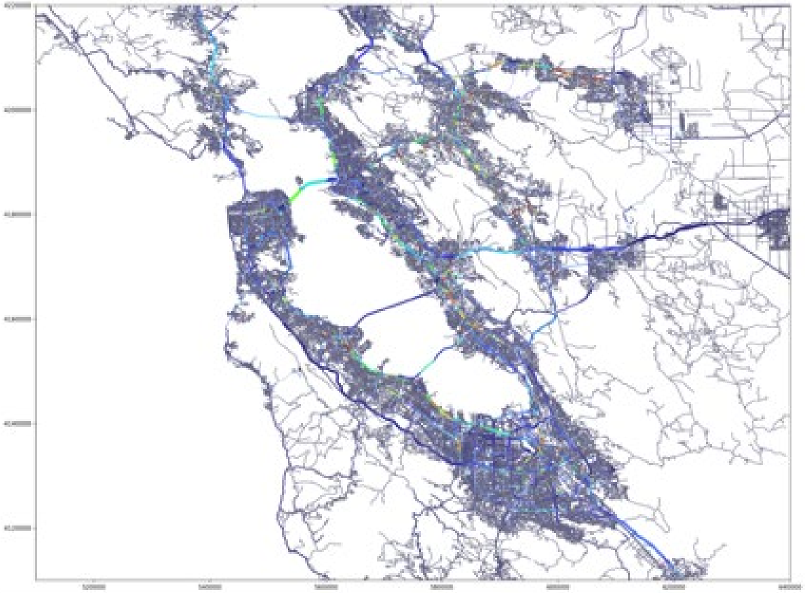
MOBILITI RESULTS WITH EXPANDED SF DEMAND MODEL

Demand 22M ODS
Network Size 2M links, 1 M nodes
Run Time 1 minute





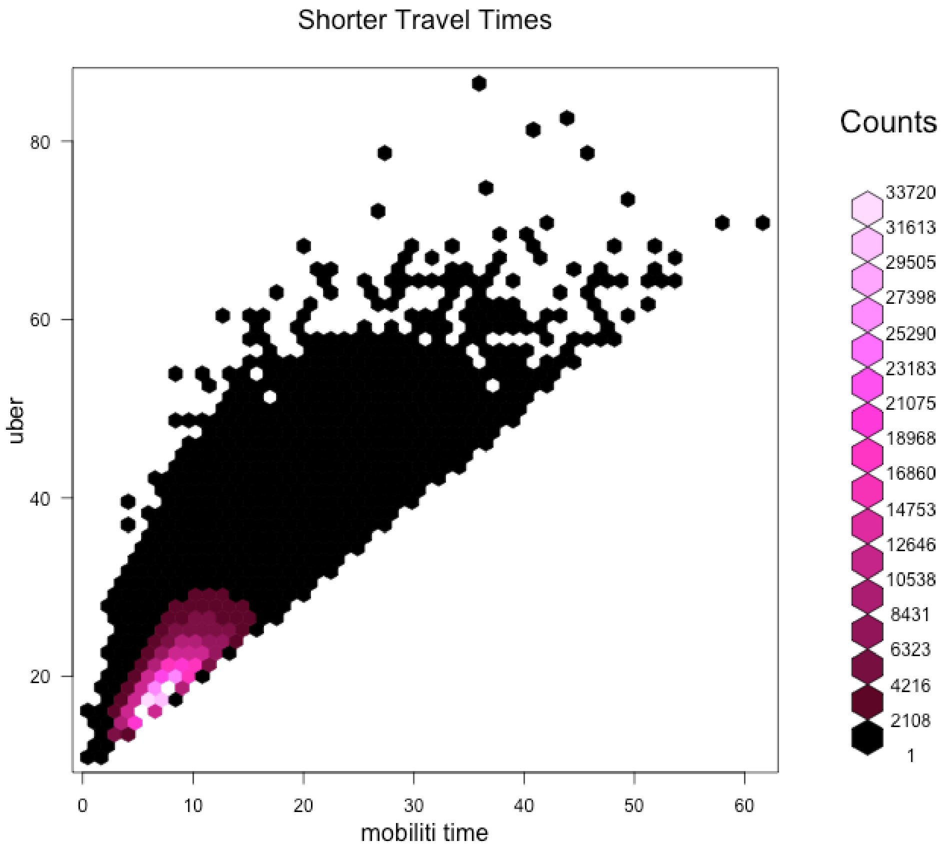
Flow Rate

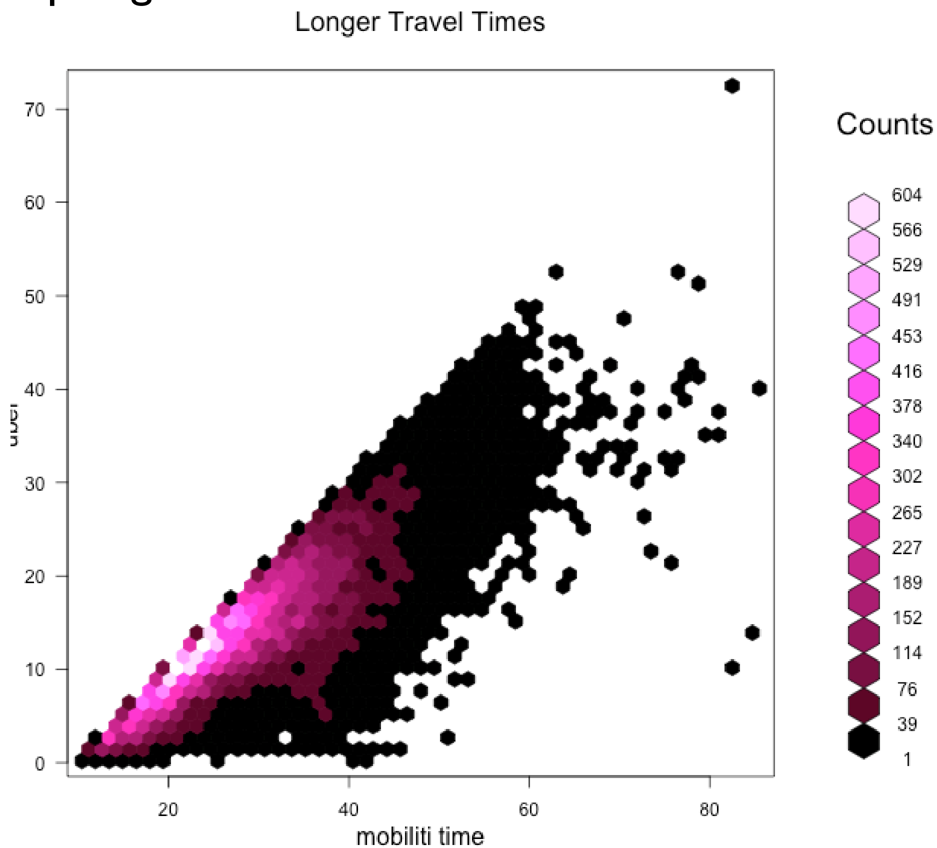


Congestion Delay

UBER MOVEMENT VALIDATION OF SIMULATION

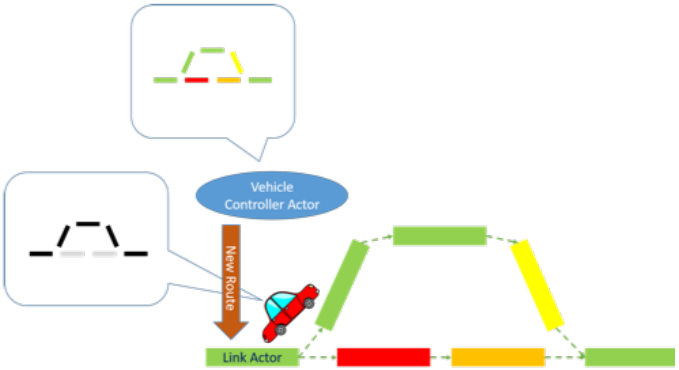
22 Million Trip Legs





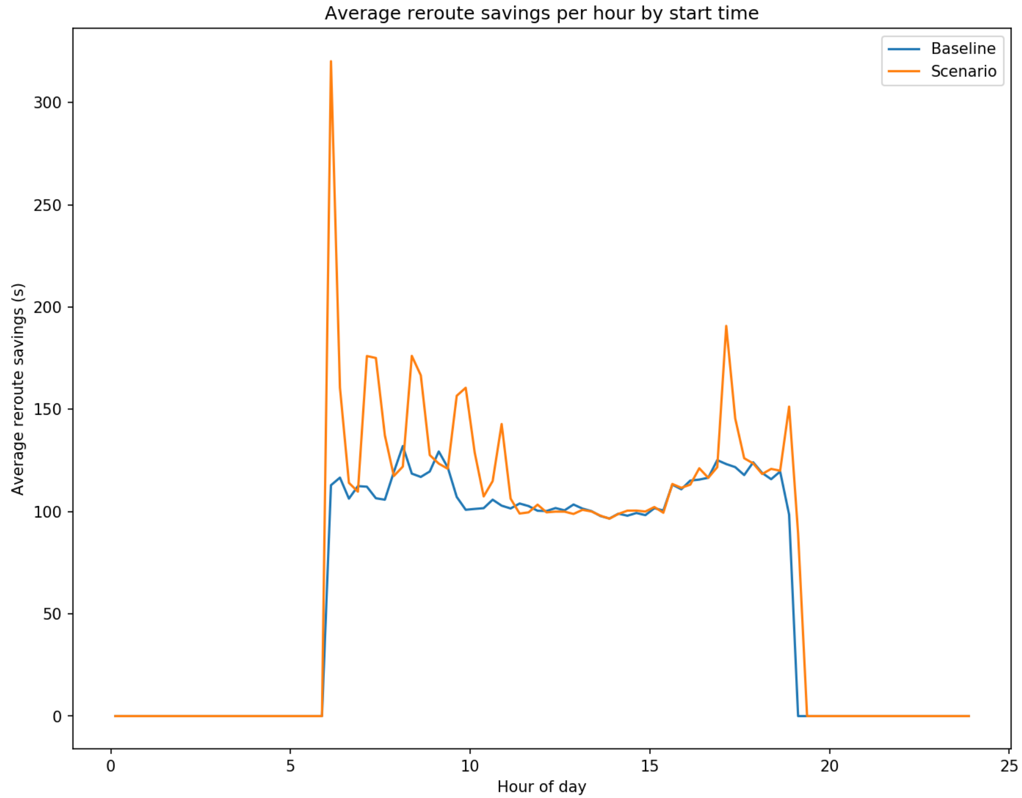
Travel time difference beyond 10 minute difference
Map anomalies or demand anomalies being investigated
OSM map currently being replaced by professional grade map

IMPACTS OF SYSTEM LEVEL DYNAMIC ROUTING



Example Event Based Reroute

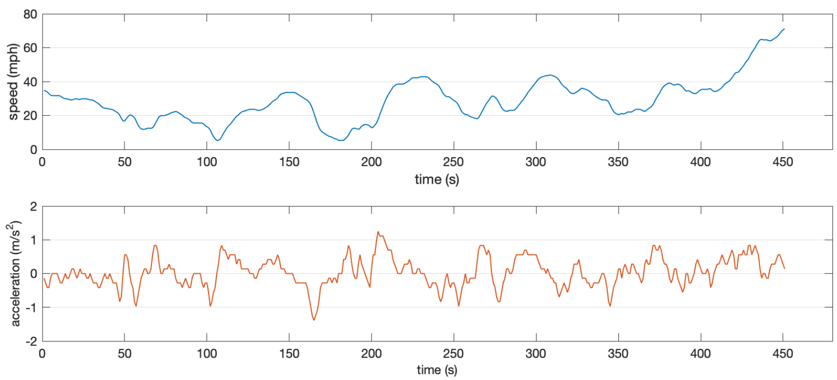
Average Reroute Time Saved per hour



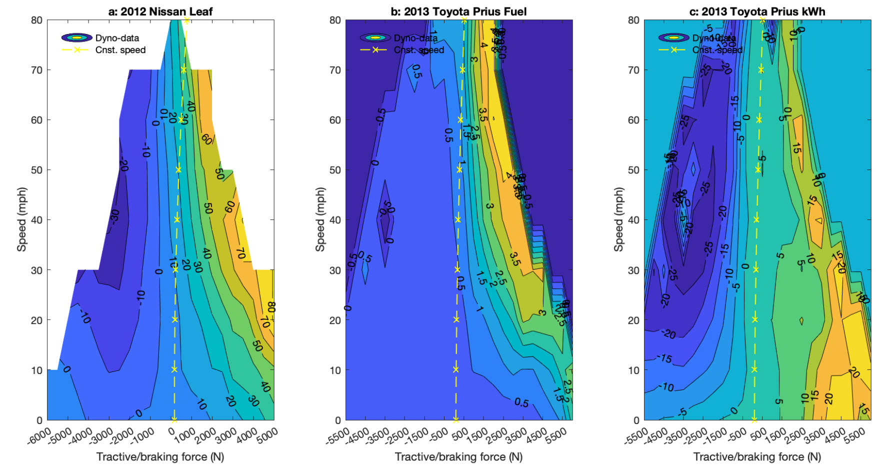
Ten million minutes of travel time, 64 thousand gallons of fuel (across 25% of vehicles), and $2.24 million productivity loss were saved due to dynamic rerouting, at the cost of increasing total distance by 368 thousand extra vehicle kilometers.

ENERGY CONSUMPTION ESTIMATES FROM REAL-WORLD MOBILE DEVICE DATA

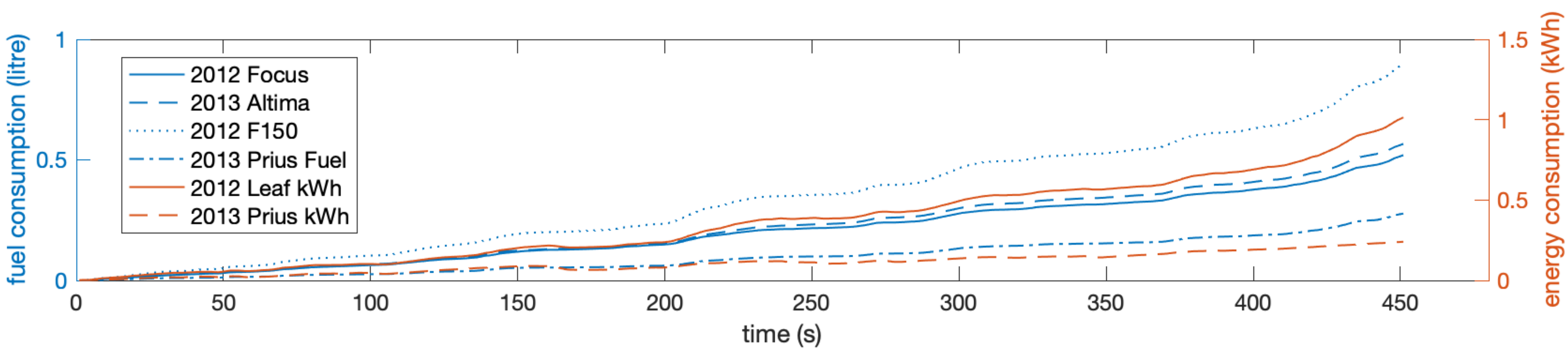
Sample Trajectory in Congestion



ML Derived Fuel and Energy Consumption Rates for Plug-In Hybrid Vehicles from ANL D3 Datasets

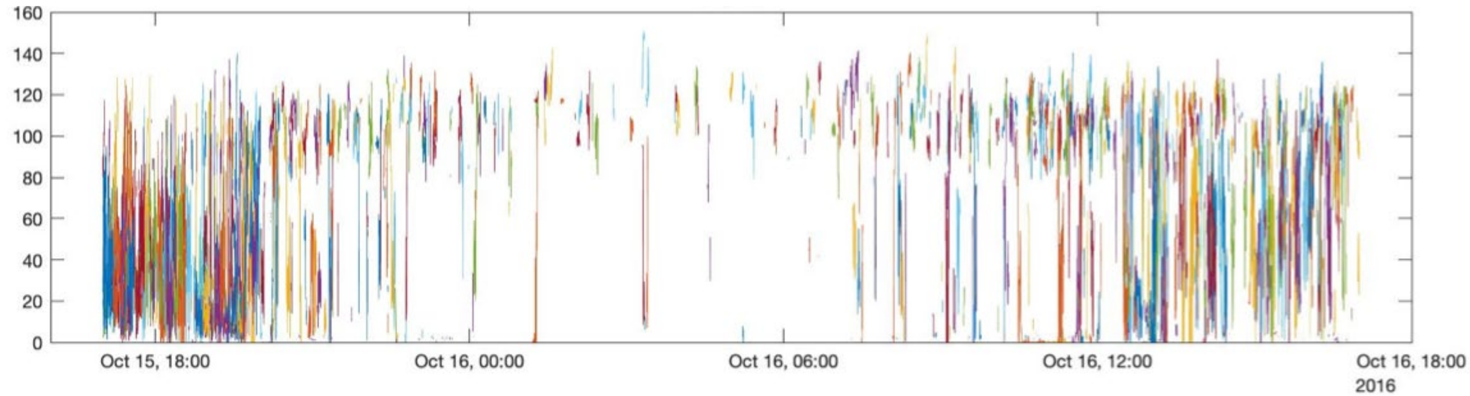


Accumulative Energy and Fuel Consumption for Sample Trajectory

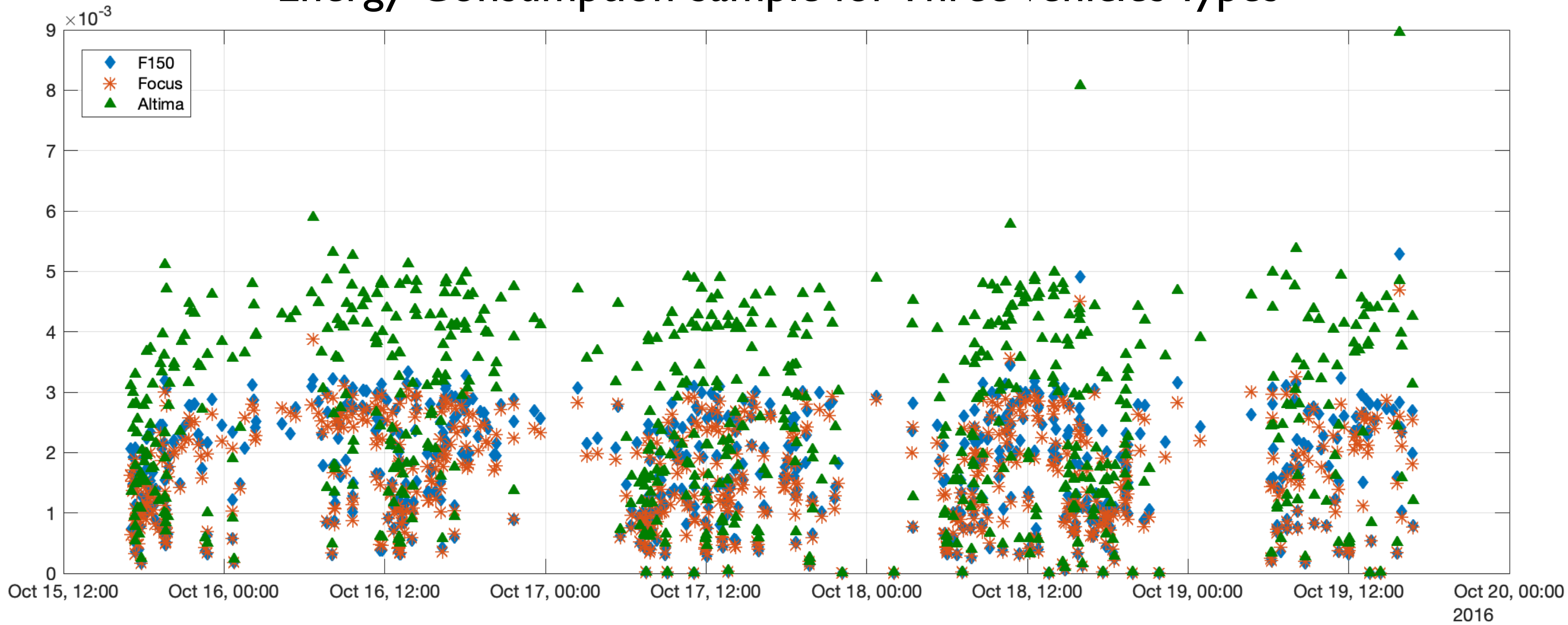


ML MODELS FOR ENERGY CONSUMPTION RATE FOR SAMPLE TRAJECTORIES ON I210

Trajectories from Mobile Devices on I210



Energy Consumption Sample for Three Vehicles Types

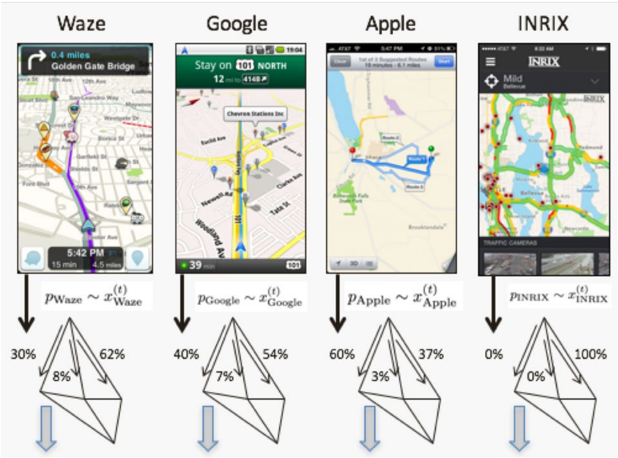


RESPONSE TO PREVIOUS YEAR COMMENTS

Comment : The problem being solved is critical to the type of simulations needed for transportation planning. Reducing computation time is critical if these and other models are going to be useful.

We thank the reviewers for the positive comments about the impact this effort can have on the goals of DOE.

Uncoordinated active control



Comment : Project Team is just scratching the surface, but that team has to think of "what the end game" is for analysis.

The impact of HPC has great promise to change the way planners approach transportation planning. We have made significant progress in the first phase by leveraging existing tools in the super computing community. We already have some active control on our roadways. We hope to provide the capability to ***design active control strategies by routing for energy reduction across the full fleet of future connected vehicles.*** *Emergency management planning could benefit greatly from metropolitan scale simulations that can be run for large numbers of scenarios with this magnitude of reduction in computation time.*

COLLABORATION AND COORDINATION

National Laboratories : HPC Modeling

BERKELEY LAB

Pacific Northwest NATIONAL LABORATORY

Argonne NATIONAL LABORATORY

HPC4 Mobility

Government and Academia : Infrastructure Data

UC Berkeley | ITS/PATH

Connected Corridors Program

SF CHAMP

CITY OF SAN JOSE CAPITAL OF SILICON VALLEY

Caltrans

TTI

Industry : Mobility Data

Uber

here

CHALLENGES AND PROPOSED FUTURE RESEARCH

- Use of mobile device data as virtual sensors to expand geospatial extent of sensing capabilities for government agencies
- Integration of additional real-world sensors – eg. weather
- Understanding how to integrate learned link dynamics into existing simulation while maintaining reduced computational time
- Validation with other simulation efforts – TTI and Texas A&M
- Validation with Uber speed data
- Validation of data driven energy estimates from mobile device trajectories

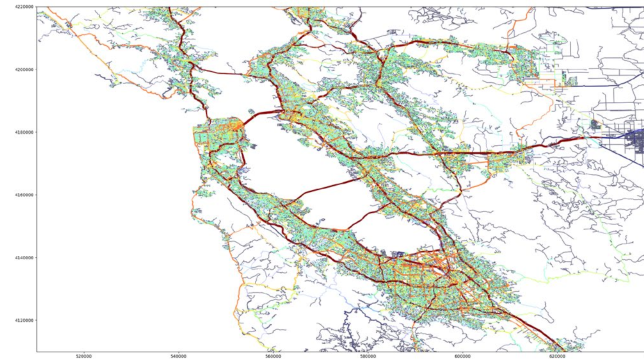
Next Phase : Surrogate ML models from Mobiliti results
Open sourcing of Mobiliti for multiple cities
Development of active control algorithms for connected vehicles

Any proposed future work is subject to change based on funding levels.

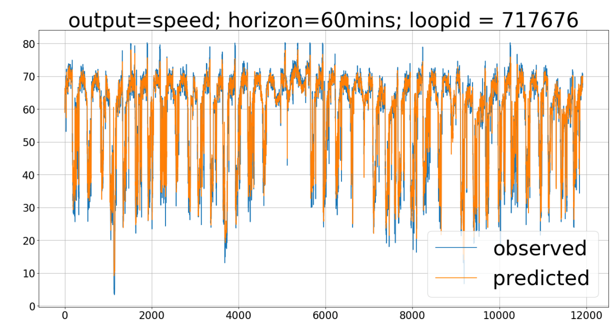
SUMMARY

- Expanded Mobiliti simulation capability to model 22M OD integrated micro analysis zones
- Developed active control mechanism in Mobiliti
- Integrated Traffic Assignment models into the framework for comparison and validation
- Validating model with Uber travel time data
- Introduced well performing ML models for capturing traffic dynamics using loop sensors
- Developed data driven energy models from mobile device trajectories for 7 vehicle types

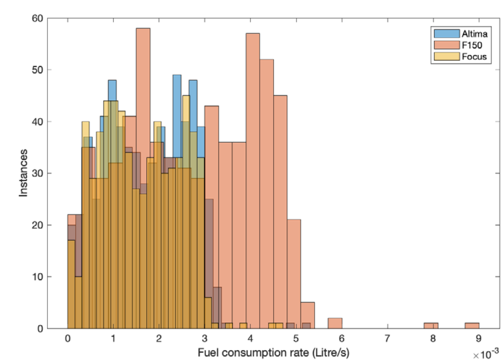
Average Link Speeds for Bay Area



DCRNN Models for Predicting Speed/Flow

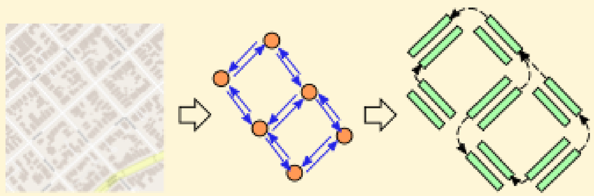


Data Driven ML Energy Models

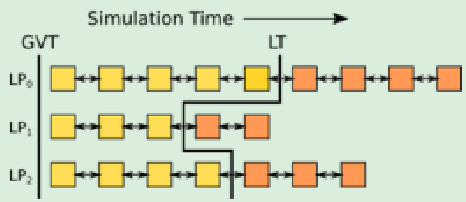


TECHNICAL BACKUP SLIDES

MOBILITI LAYERED ARCHITECTURE



Mobiliti: provides the domain-specific logic that defines the actors and events of a traffic system, and determines the parallel domain decomposition mapping actors to ranks

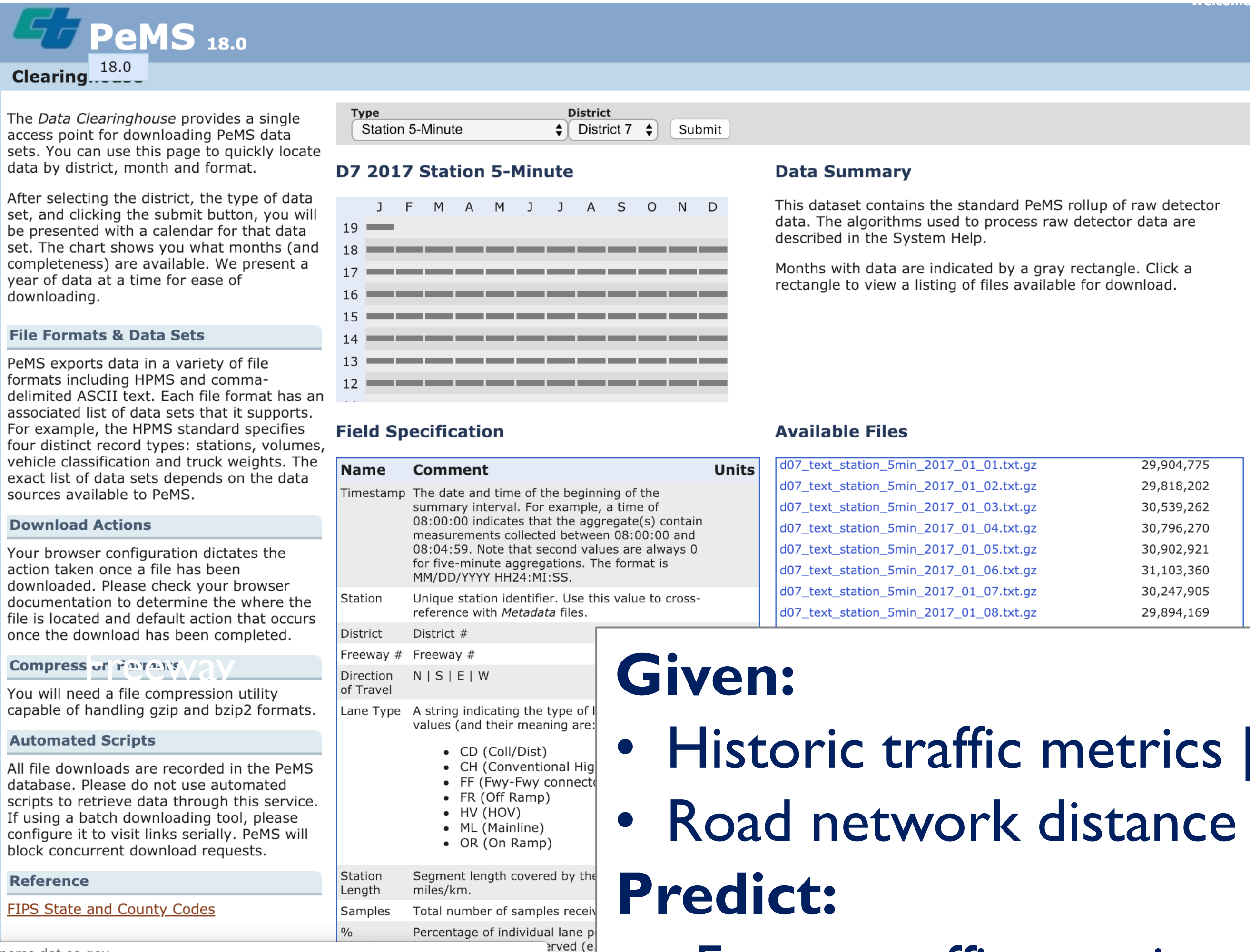


Devastator: implements Jefferson's Time Warp optimistic parallel discrete event protocol [1] to handle event scheduling, execution, rollback, commit, and global virtual time

GASNet-Ex: provides high-performance inter-process communications across distributed memory, in particular for small active messages

[1] David R. Jefferson. 1985. Virtual time. ACM Trans. Program. Lang. Syst. 7, 3 (July 1985), 404-425

USING ML FOR PREDICTING TRAFFIC METRICS



Data includes:
- Timestamp
- Loop IDs
- District
- Freeway name
- Freeway direction
- Total flow
- Average speed

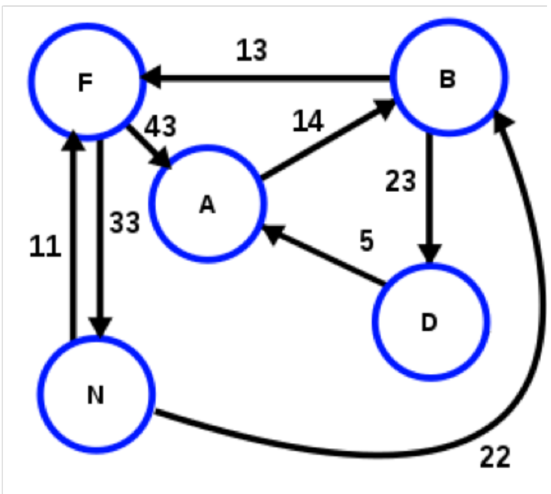
- ~18K sensors

Given:
- Historic traffic metrics [speed, flow]
- Road network distance and connectivity

Predict:

Future traffic metrics

GRAPH REPRESENTATION OF ROAD NETWORK



Transportation network as graph

- V = Vertices (sensors)
- E = Edges (roads)
- A = Weighted adjacency matrix

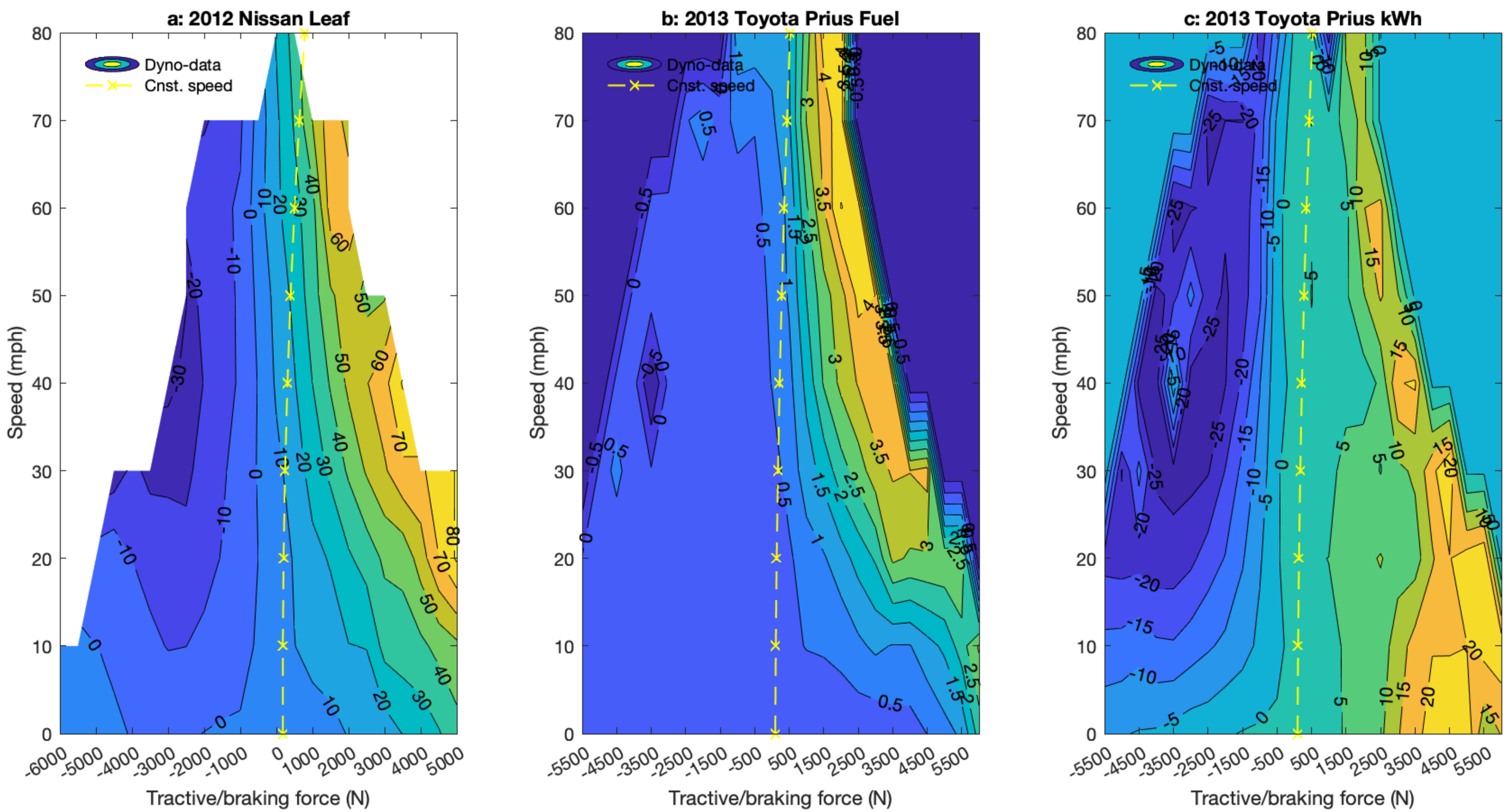
  (A function of the road network distance)

$$A_{ij} = \exp\left(-\frac{\mathrm{dist}_{\mathrm{net}}(v_i, v_j)^2}{\sigma^2}\right) \text{ if } \mathrm{dist}_{\mathrm{net}}(v_i, v_j) \leq \kappa$$

$\mathrm{dist}_{\mathrm{net}}(v_i, v_j)$: road network distance from v_i to v_j,

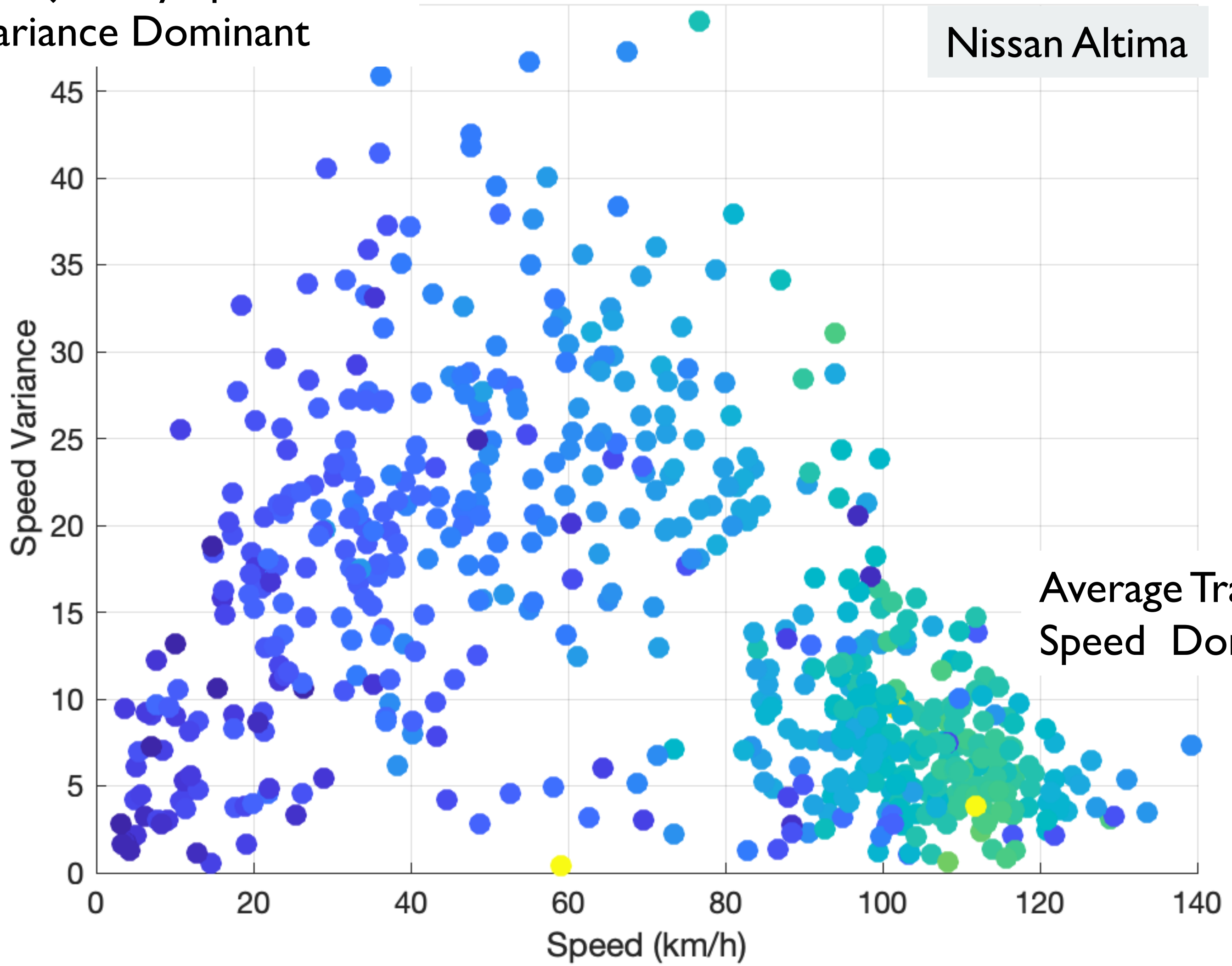
κ: threshold to ensure sparsity, σ^2 variance of all pairwise road network distances

LSTM MODELS FOR ENERGY CONSUMPTION PLUG-IN HYBRID ELECTRIC VEHICLES



TRAJECTORY FUEL CONSUMPTION RATES WRT TRAJECTORY AVERAGE SPEED AND SPEED VARIANCE

Average Trajectory Speed and Speed Variance Dominant



Average Trajectory Speed Dominant